基于产品评论的消费者情感波动分析模型构建及实证研究*

林园园　战洪飞　余军合　李长江　张　凡
(宁波大学机械工程与力学学院　宁波 315211)

摘要:【目的】基于网络产品评论信息构建消费者的情感波动分析模型。【方法】该分析模型以文本挖掘技术理论为基础,运用产品评论挖掘技术和情感分析技术,在情感分析的同时充分考虑不同类型连词对句子情感倾向性的影响,并采用相应的权值计算方法。【结果】从京东和中关村在线抓取某款手机从 2013 年 11 月到 2015 年 1 月这段时间内产品评论信息并进行分析,验证了该模型的有效性。【局限】在分析消费者情感波动主要影响因素方面,该分析模型主要考虑相邻时间段内产品特征词个数的变化以及产品特征词在评论信息中出现次数的变化这两个维度,其他维度并未涉及。【结论】该模型有效地分析了消费者在一段时间内的情感波动趋势,以及产生情感波动的因素,能够为企业决策提供一定的参考。

关键词: 数据分析　文本挖掘　情感波动分析模型
分类号: TP391.1　G202

1 引 言

随着互联网的不断发展和深入人们的生活,互联网思维逐渐渗透到生活的方方面面,越来越多的人们开始借助网络来表达他们对某件事物的观点和情感态度。特别是随着电子商务的兴起,在互联网上购买商品对大多数用户来说变得不那么陌生,用户开始在网站、论坛、讨论组、微博、微信等平台发表他们对某产品的功能、性能和服务等方面的态度和意见,在这些用户评论中蕴含着可供企业做出决策的重要信息,然而面对众多的网络评论信息,很难准确地从中获得所需要的信息,所以需要运用技术手段来解决这些问题。在目前产品评论挖掘的研究中,基本上都是利用文本挖掘、信息检索、信息过滤和自然语言处理等技术。通过这种方式可以在较短时间内获得有价值的信息。

Xu 等[1]将文本挖掘技术运用到企业的竞争情报中,通过构建模型来抽取竞争产品之间的关系。Rodrigues 等[2]提出一种方法,基于 SentiHealth 癌症患者在社交网络上的文本信息来分析其心理状况,同时自动分析癌症患者的社会岗位。Salehan 等[3]提出一种情感挖掘的大数据分析方法以分析消费者在线评论信息对消费者购买产品的影响。Deng 等[4]认为虚假餐馆评论误导消费者的选择,提出一种基于情感分析的算法来检测网上餐馆虚假评论。Wallace 等[5]基于 f-LDA 模型按用户关心的各个方面抽取医生在线评论进行情感分类。翟东升等[6]将文本挖掘技术运用到产品的竞争优势分析上,通过计算两种竞争产品所关注的产品特征和产品的极性值,发现竞争对手的优势及自身需要改进的地方。施国良等[7]把文本挖掘技术运用到对同一种商品在不同购物网站的一致性研究上面。王伟等[8]则基于文本挖掘技术研究在线评论信息对用

通讯作者: 战洪飞, ORCID: 0000-0001-5029-2222, E-mail: zhanhongfei@nbu.edu.cn。

*本文系浙江省自然科学基金项目“中小企业业务问题及其解决方案的建模理论与管理方法研究”(项目编号: LY14G010003)、国家星火计划项目“农业业务问题的大数据推荐技术应用与示范”(项目编号: 2015GA701011)和国家自然科学基金项目“面向产业集群广义制造系统的知识模块化机理研究”(项目编号: 71671097)的研究成果之一。

户购买意愿的影响，通过构建模型分析产品特征评价与用户购买意愿的关系从而识别出重要的产品特征。吴丽华等[9]则利用文本挖掘技术实现 IT 产品特征和相关情感倾向的挖掘发现。田雪筠[10]根据产品评论信息设计一种算法来获取企业的竞争情报信息。吴丽云等[11]基于文本挖掘技术，以大众点评网为研究对象，根据评论内容研究消费者的行为。从以上研究可以看出，大多数学者在进行产品评论挖掘研究时只是从不同的角度分析问题，在这些研究中只分析了消费者所关注的产品特征和对所关注的产品特征的情感倾性向分析，并没有从整体层面考虑消费者在一段时间内的情感波动趋势。还有一些学者如：黄卫东等[12]提出一种基于概率潜在语义分析的网络舆情话题情感分析方法，分析了不同时间段内网络舆情话题子话题随时间演化过程。倪瑜泽等[13]提出一种基于用户评论的潜在演化需求发现方法，该方法有效地辅助需求分析师进行演化需求的抽取。何天翔等[14]使用 DTM 模型对微博数据进行时间分片，然后进行情感演化分析从而找到舆情变化的时间点。李超雄等[15]提出一种动态主题情感混合模型，该模型相比其他模型具有更高的情感分类准确率并且可以进行微博主题情感演化分析。尽管这些学者在整体层面对网络信息进行了分析，但只是进行了情感演化分析和评论演化分析，并未进一步挖掘什么因素导致消费者的情感发生变化。所以本文正是运用评论挖掘等技术，构建消费者情感波动分析模型，分析客户在某段时间内对产品的情感波动情况，同时找出产生情感波动的因素，帮助企业掌握消费者对产品的情感状态趋势，辅助企业做出决策。

本文将产品评论挖掘技术和情感分析技术应用于消费者对产品的情感波动分析领域，构建消费者对产品的情感波动分析模型。该模型不仅分析了消费者所关注的产品特征以及对产品特征评价的情感倾向，还分析了在一段时间内消费者对产品的情感波动趋势以及引起消费者情感产生波动的因素。

2 研究模型的构建

网络上消费者对产品的评论一般都包含消费者使用产品后内心的真实想法，反映出他们内心的情感状态。许多企业已在互联网上开辟了专门的评论专区、论坛、社区等供消费者表达自己的想法，企业则可以了解消费者对产品的看法等，但是企业只是了解消费者关注的产品特征和消费者对产品某些特征的情感倾向等，并没有从整体层面关注消费者的情感波动情况以及进一步了解影响消费者情感波动的因素。企业可以根据情感波动趋势图及波动产生的因素做出相应的应对措施，同时还可以根据情感波动趋势图结合自身的销售数据趋势图做对比，找出内在联系供企业做参考。所以本文从整体层面考虑消费者的情感波动情况，构建情感波动分析模型，如图 1 所示。根据产品评论挖掘的一般步骤，情感波动分析模型可分为 4 个层面：数据抓取层、数据清理层、产品评论挖掘层、情感波动分析层。

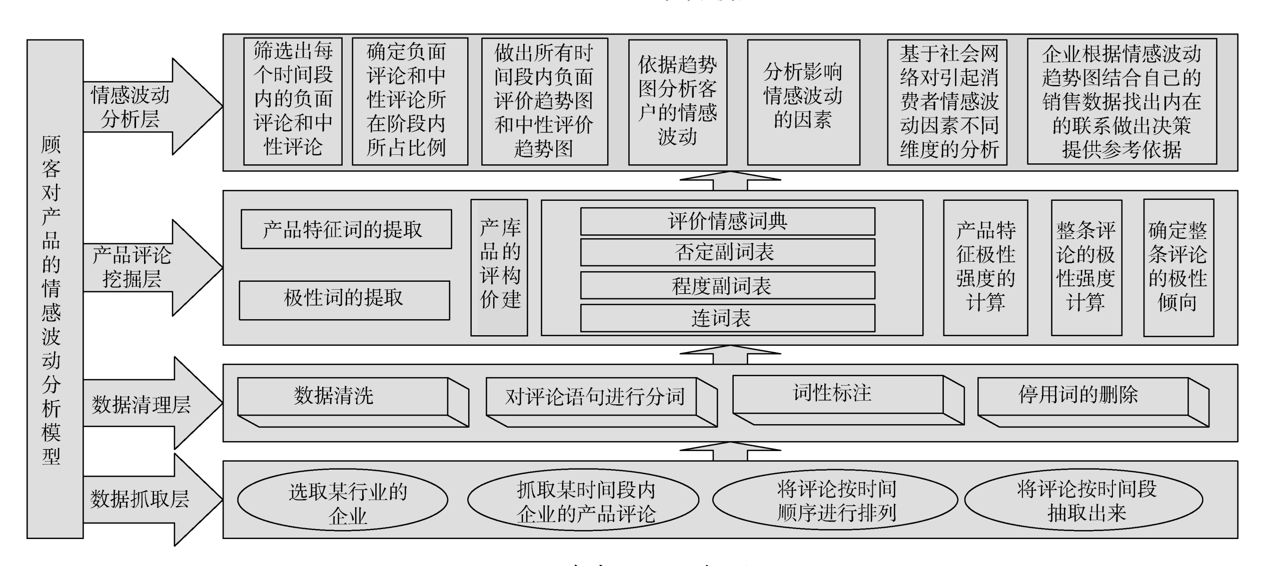


图 1 情感波动分析模型

(1) 情感波动分析模型从底层的数据抓取层开始，首先运用网络爬虫爬取产品的评论信息，产品的评论信息包括评论的时间和评论的内容，把产品评论信息存放于 Excel 表格中并且按时间顺序排列。

(2) 数据清理层用 R 语言编写程序对这些评论信息进行数据清理、分词、词性标注和停用词的删除。

(3) 产品评论挖掘阶段，主要是对产品特征词的提取、极性词的提取、产品评价库的构造、情感词典的构造、产品特征极性强度的计算和整条评论信息的极性计算。为了能够更加全面地分析消费者的情感状态，在考虑程度副词和否定词对评价短语的情感倾向性影响后，也将连词对评价语句的情感倾向性影响考虑进来，同时给出相应评价短语的权值计算方法。

(4) 情感波动分析层，利用提出的情感波动分析方法对整个时间段内消费者对产品的情感波动情况进行分析。

3　消费者情感波动分析模型具体研究过程

消费者情感波动分析模型具体研究过程包括：数据的抓取和预处理、产品评论挖掘过程、情感波动分析过程。具体流程如图 2 所示：

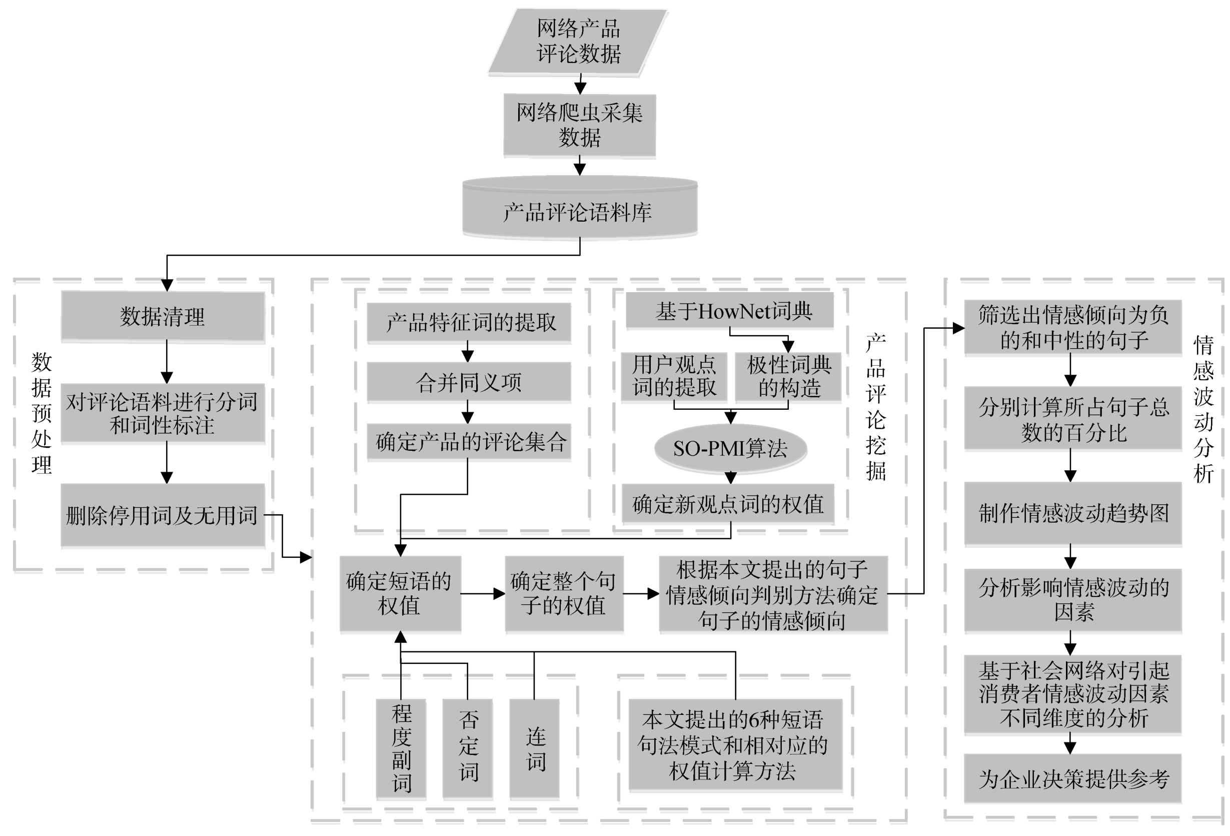


图 2　情感波动分析过程流程图

3.1　数据的抓取和预处理

针对本文所需要的数据，编写网络爬虫程序爬取产品评论信息，构建产品评论语料库。具体包括：评论的时间和评论的内容。然而，通过网络爬虫采集到的评论信息其中含有大量与用户观点态度无关的信息，这些信息最终会影响产品评论挖掘的结果，所以首先需要对网络爬虫采集到的信息进行处理。处理的过程包含三个方面：数据的清理、分词、删除停用词。

(1) 数据清理

在消费者对产品的评论信息中出现的消费者喜欢用的口语化词语、网络新词、图片和音频等非文本信息，还有一些与产品特征和用户观点没有关系的评论语句等，这些信息可以视为噪声，可以把这些评论语句删除，最后提取对产品有意义的用户评论。

(2) 分词处理

经过数据清理后的评论语句暂时还无法直接应用情感分析的方法进行处理，需要将整个没有结构的文本转化为结构化的数据形式，而最常用的方法就是将文本表示为词汇的形式。实现这一步需要首先进行分词，分词的好坏直接决定了计算机对文本语义分析的准确性。本文借助 R 软件中的分词工具 Rwordseg 完成中文分词。Rwordseg 使用 rJava 调用 Java 分词工具 Ansj，Ansj 是基于中国科学院计算技术研究所 ICTCALS 中文分词算法的开源工具，采用隐马尔科夫模型(Hidden Markov Model, HMM)，具有很好的分词效果[16]。

(3) 词性标注

在进行分词的时候已经对每个词语进行了词性标注，通过词性标注可以判定每个词的语法范畴，是属于观点词、特征词、还是程度副词，从而提取出产品特征和识别用户观点态度。本文词性标注采用统计的方法，基于隐马尔科夫模型，对文本进行分词得到词序列{W_1，W_2，…}，词性序列{t_1，t_2，…}为隐含着的状态序列，然后基于 Viterbi 算法求解统计词性转移矩阵[a_{ij}]和词性到词语的输出矩阵，求解的过程实际上是求可能性最大的状态序列[17]。

(4) 删除停用词及无用词

在实际应用中，文本中很多词是与内容无关的，比如“呢”、“啊”、“他们”等，因此可将这些与内容无关的词(比如代词、介词、拟声词等)从文本中去除，然后通过对文档中词频的统计，将文本中出现频率过高或过低的并且对于内容没有太多贡献的词语从文本中去除，这样会提高产品特征分析的效果和效率。本文通过 R 语言编写程序读取停用词表(1 893 个)对文档进行处理。

3.2　产品评论挖掘的过程

经过数据的抓取和预处理后，进入产品的评论挖掘阶段。产品的评论挖掘过程主要包括产品特征的提取、极性强度的确定、极性强度词典的构造和评价短语的情感倾向性计算。

(1) 产品特征的提取

在进行产品评论挖掘的过程中，首先需要对产品特征词进行提取，从用户评论中抽取出用户所关心的产品特征，比如手机的产品特征包括：屏幕、电池、配置等。经过词性标注后提取全部名词，通过筛选去除和产品特征无关的词汇，最后得到用户关心的产品特征集合。

(2) 极性强度的确定

极性词是人们用来表达自己主观感受的观点词，有喜、怒、哀、乐等。当人们对某一事物的好坏做出评价时，往往通过使用这类词来表达自己的观点态度。极性词一般包括形容词、动词和部分的名词。首先在 2007 年知网发布的“情感分析用词语集(beta 版)”中挑选与待分析产品紧密相关的词汇构成用户极性词典[18]。这些词汇有代表褒义的，也有代表贬义的，并且这些词汇是最基本的包含强烈褒贬色彩的词语，是大量语料中总结出来的最常用的极性词语。通过采用 Turney 等[19]的 SO-PMI 算法计算新词与这些基准词在语料库中的共现概率，确定新词的褒贬义倾向及强度。

(3) 极性强度词典的构造

①程度副词

程度副词会影响情感词的强弱，当程度副词改变的时候，情感词的极性也发生改变，所以在对句子进行极性分析的时候，要将程度副词影响程度考虑在内。程度副词可分为相对程度副词和绝对程度副词。相对程度副词和绝对程度副词又可分为极量、高量、中量和低量 4 种程度，不同等级的程度副词赋予不同的权重[20]，分别为：极量(1.5)、高量(1.2)、中量(0.9)、低量(0.6)。常用的相对程度副词和绝对程度副词，如表 1 所示：

表 1　程度副词表

程度	相对程度副词	绝对程度副词
极量	最、最为、无比	太、极、极为、极其、极度、过、过于、过分、分外、万分
高量	更、更加、更为、越、越发、备加、愈、愈加、越加、格外、益发、愈益	很、挺、怪、老、非常、特别、相当、十分、好、颇、颇为、异常、深为、蛮、够、多、多么、特、尤其、无比、尤为
中量	较、比较、较比、还	不太、不很、不甚
低量	稍、稍稍、稍微、稍许、略、略微、多少	有点、有些

②连词表

在对评论语句进行极性分析的过程中，往往会遇到很多的连词，如并列连词、转折连词、递进连词、让步连词等。在这些连词当中，最多的即是并列连词、转折连词、递进连词，它们会影响整个句子情感表达的侧重。笔者总结出常用的并列连词、转折连词和递进连词，如表 2 所示。

1) 如果评论句中出现转折连词，则在连词前后的语义极性是相反的。

2) 如果评论句中出现并列连词，则该连词之前和之后的语义极性是相同的。

表 2　连词表

类型	连词
转折	却、虽然、但是、然而、而、偏偏、只是、不过、至于、致、不料、岂知
递进	不但、不仅、而且、何况、并、且
并列	和、跟、与、既、同、及、而、况、况且、何况、乃至

3) 如果评论句中出现表示递进的关系连词, 则该词之后的语义极性比之前的语义极性更强。

③否定词

否定词在对计算评价短语的情感倾向性权值时有很大的影响, 如不、不是等。这些否定词在修饰情感词时起到反向的作用, 但是一般遇到否定词修饰情感词的时候并不是完全起反向的作用, 只是对情感词的情感起到一个弱化的作用。所以在有否定词修饰的情感词时, 在情感词的前面乘以–0.7。经过筛选得到最常用的否定词, 如表 3 所示:

表 3　否定词表

否定词						
不	非	别	欠	没	未	缺
不怎么	不觉得	毫不	算不上	不会	无法	木有
说不上	并未	不可能	没有	不太	尚未	难以
不是	不必	决不	从未	不能	不足	欠妥

(4) 评价短语的情感倾向性权值计算

评价短语的情感倾向性权值计算的准确度对于判断整个句子的情感倾向性有重要的影响, 所以本文在魏慧玲[21]给出的 5 种情况下又增加一种新的情况, 同时给出计算权值的方法。这样可以更进一步提高评价短语情感倾向性权值计算的精确度, 对产品评论挖掘效果会有很大的提升。这 6 种情况分别为: 情感词; 否定词+情感词; 程度副词+情感词; 程度副词+否定词+情感词; 否定词+程度副词+情感词; 情感词+连词+情感词。

对于以上 6 种情况, 给出对应权值的计算方法, 前 5 种分别是: 直接计算情感词的权值; 情感词权值×(–0.7); 程度副词权值×情感词权值; (–0.7)×程度副词权值×情感词权值; (–0.7)×程度副词权值×情感词权值。最后一种比较复杂, 可以进一步细分:

①如果是转折连词: 按转折连词后面的情感词的权值进行计算。

②如果是并列连词: 其中一个情感词的权值为 a, 另一个权值为 b。那么评价短语的情感倾向性权值为 $c = \frac{a+b}{2}$。

③如果是递进连词: 其中一个情感词的权值为 a, 另一个权值为 b。那么评价短语的情感倾向性权值为 $c = \frac{a+b}{2} \times 1.5$。

最后运用公式(1)确定整个句子的情感倾向:

$$\text{Score} = \frac{\sum_{i=1}^{m} P_m + \sum_{j=1}^{n} P_n}{m+n} \tag{1}$$

其中, P_m 代表特征词的极性权值为正的集合, P_n 代表特征词的极性权值为负的集合, m 代表一个句子中极性值为正的特征词的个数, n 代表极性值为负的特征词的个数。按照给出的 6 种权值的计算方法计算 P_m 和 P_n 的权值。

计算完成一个句子的权值后需判定一个句子的极性为正还是为负, 判别条件如下:

$$\text{class} = \begin{cases} c_1 & \text{Score} > 0 \\ c_2 & \text{Score} = 0 \\ c_3 & \text{Score} < 0 \end{cases} \tag{2}$$

其中, c_1 表示极性为正, c_2 表示极性为中性, c_3 表示极性为负。

3.3　网络产品的情感波动分析过程与方法

企业之间的竞争越来越激烈, 而用户在企业的竞争过程中处于核心地位。谁能获得更多的用户, 谁就能在企业的竞争过程中处于有利地位。如何使自己的产品更吸引用户, 是企业值得思考的问题。在吸引用户基础之上需要了解用户对产品的心理状况, 因此掌握用户对产品的情感波动情况, 找出波动的主要因素并能够做出相应的决策以吸引用户进而提升销售业绩, 对提升企业的竞争力具有重要的意义, 企业平时关注最多的就是通过消费者对产品的评价信息来了解产品在哪些地方做得不好或者哪些地方做得还可以。本文根据单晓红等[22]提出的 9 种语法模式, 筛选所需要的句子。要筛选掉的句子有两种: 不包含特征词集合里面的特征词; 不符合这 9 种语法模式的句子。最后找出消费者情感倾向极性值为负和情感倾向为中性的句子作分析。对消费者的情感波动分析主要从以下方面进行: 制作情感波动趋势图; 分析影响情感波动的主要因素; 基于社会网络对引起消费者情感波动因素不同维度的分析。

(1) 制作情感波动趋势图

根据筛选后的产品评论语句集合, 计算每个句子的情感倾向, 以每月为单位将其中的情感倾向性权值为负的、情感倾向性权值为中性的句子抽取出来, 分别计算它们所占句子总数的百分比。计算公式如下。

$$p = \frac{q}{n} \times 100\% \quad (3)$$

其中，n 代表每个月的句子的总数，q 代表情感倾向性权值为负的或者为中性的句子总数。

(2) 分析影响情感波动的主要因素

分析情感波动的主要因素主要是分析负面情感波动，首先根据情感波动趋势图的走势，分别把上升段、下降段、平稳段的几个月做一个集合，分析引起情感波动主要因素的方法是分别抽取在情感波动趋势图上升段、下降段、平稳段这几个时间段内消费者对产品特征负面评论的特征词，分别为：上升段负面特征词集合 $a=\{a_1, a_2, a_3 \cdots a_n\}$，下降段负面特征词集合 $b=\{b_1, b_2, b_3 \cdots b_n\}$，平稳段负面特证词集合 $c=\{c_1, c_2, c_3 \cdots c_n\}$，这些特征词集合是影响消费者情感波动的体现，然后对比集合 a，b，c 中产品特征词的个数是否变化很大，当集合中产品特征词的个数变化很大的时候，可以考虑是否主要是因为产品特征词个数的变化导致负面评论比例的上升，从而导致情感发生波动，同时找出增加的产品特征词，这些增加的产品特征词是影响消费者情感波动的主要因素。当相邻时间段产品特征词个数变化不大的时候，分别对比相邻时间段内相同产品特征词在评论信息中出现的次数是否发生很大变化，这些产品特征词在评论信息中次数的增加导致负面评论比例的上升，从而导致消费者情感波动发生变化。

(3) 基于社会网络对引起消费者情感波动因素不同维度的分析

经过对情感波动主要因素的分析，生产厂家很容易知道消费者对产品的哪些特征不满意，但是生产厂家并不知道消费者从哪些维度去评价产品特征的好与不好，所以本文运用关键词共现分析方法，其原理是对一组词两两统计其在同一篇文献中出现的次数，以此为基础对这些词进行分层聚类，通过构建关键词共现矩阵，可以表明哪两个词之间具有一定的内在联系，同时基于 Ucinet 对共现矩阵进行可视化分析，直观展现出关键词之间的联系。

4 分析模型的实例验证

4.1 网络信息采集

某款手机的评论数据来自于中关村在线(http://www.zol.com.cn)和京东商城(http://www.jd.com)，采集时间从 2013 年 11 月 1 日到 2015 年 1 月 31 日，共采集到 14 890 条信息。采集的评论数据包括评论的时间、评论的内容，根据此信息构建产品评论语料库。示例如表 4 所示：

表 4 产品评论语料库

时间	评论内容
2013-11-1	屏幕分辩率高 细腻，机身薄，系统流畅，性能彪悍。
2013-11-1	配置高、屏幕大。
2013-11-1	高分辨率，音质很好，外放很棒。
2013-11-2	外观很漂亮，电池还行。

4.2 数据预处理

经过数据的清理后，共计有 12 223 条信息。利用 R 软件编写程序对这些评论语句进行分词、词性标注、删除停用词。

4.3 产品评论挖掘

(1) 产品特征的提取

首先进行特征词提取，用 R 语言编写代码提取所有的名词，再人工筛选出与所关注的产品无关的名词将其剔除掉，并合并同义特征词，如价格={报价、价钱、价位、价格}等。最终得到产品特征词的集合 F={屏幕、游戏、电池、系统、外观、配置、性能、内存、音质、价格、摄像头、处理器、铃声、硬件、音乐、闪光灯、亮度、充电器、听筒、触摸屏、耳机、相机、手感}。

(2) 评论观点极性及强度判断

①用户极性词典构造

利用 2007 年知网发布的“情感分析用词语集(beta 版)”中的中文正面情感词、负面情感词、正面评价语、负面评价语 4 个文件中选择与手机相关的词汇构成用户极性词典。其中选择 10 对基准词语{(清晰，模糊)、(流畅，卡)、(好，坏)、(没用，实用)、(强大，低级)、(漂亮，难看)、(便宜，昂贵)、(大，小)、(精细，粗糙)、(丰富，稀少)}。

②极性强度词典的构造

表 1 至表 3 列出来的程度副词、连词、否定词，这些即为构造的极性强度词典。

4.4 产品情感波动分析

(1) 绘制产品特征词词云

运用 R 语言编程绘制产品的特征词词云，从词云可以看出消费者大致关注的产品特征，字体大的表示消费者关注度高，字体小的则表示消费者关注度低，如图 3 所示。

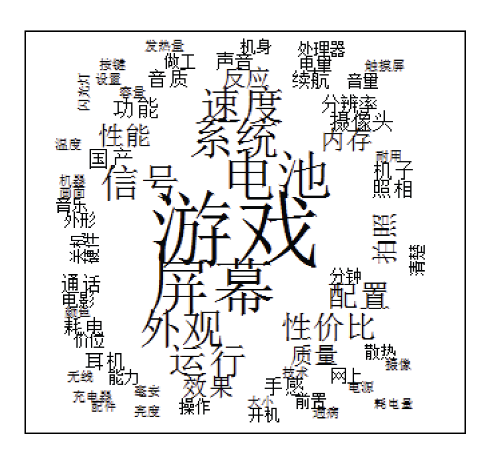


图 3　产品特征词词云

(2) 制作情感波动分析图

根据情感波动分析方法，绘制消费者中性情感波动趋势图和负面情感波动趋势图，分别如图 4 和图 5 所示。

从图 4 可以看出消费者的中性评论比例基本维持在一个相对不高的水平，最高水平在 23%左右。除个别几个月的中性评论比例在 15%左右，其他月份都在 20%左右。从整体来看中性情感波动不大。

由图 5 可以看出，从 2013 年 11 月–2014 年 1 月消费者的负面情绪比例趋于上升的趋势，2014 年 2 月–2014 年 4 月消费者的负面情绪比例趋于平稳的趋势，2014 年 5 月–2014 年 10 月消费者的负面情绪比例上升很明显，2014 年 11 月–2015 年 1 月消费者的负面情绪比例趋于下降，但总体上看消费者的负面情感还是趋于上升趋势。

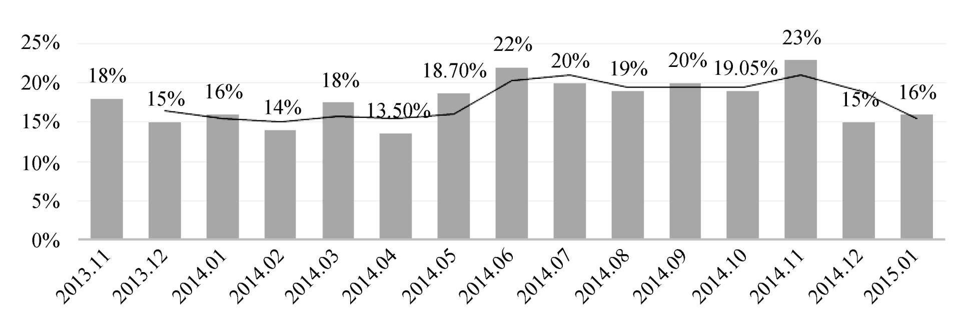


图 4　中性情感波动分析图

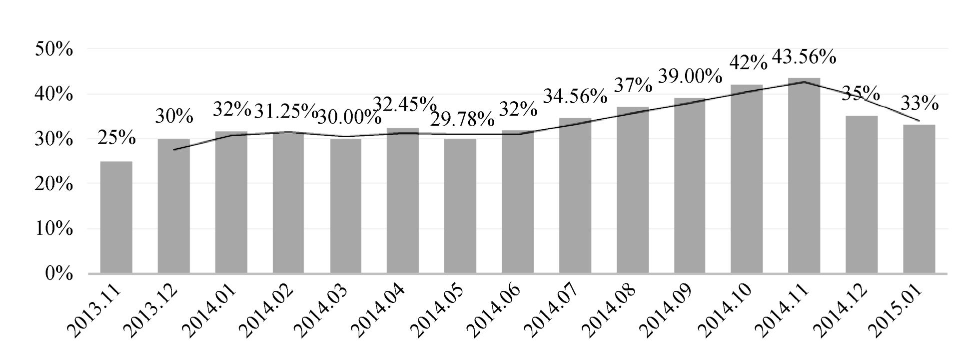


图 5　负面情感波动分析图

(3) 情感波动主要因素分析

根据负面情感波动分析图，分析不同时间段内产生这些波动的主要因素以及这些因素所占的比例，结果如图 6 所示。

从图 6 可以看出在 2013 年 11 月–2014 年 1 月这段时间内顾客的情感波动主要体现在游戏、电池、屏幕、声音、系统、外观、摄像头、信号和音频。在 2014 年 2 月–2014 年 4 月这段时间内顾客的情感波动主要体现在游戏、电池、屏幕、信号、配置、耳机、摄像头、外观、性能、系统、音量、手感、功能、音质。在 2014 年 5 月–2014 年 10 月这段时间内顾客的情感波动主要体现在游戏、屏幕、信号、电池、内存、系统、速度、声音、音乐、音质、摄像头、外观、闪光灯、功能、硬件、充电器、配置、性能和声音。在 2014 年 11 月–2015 年 1 月这段时间内顾客的情感波动主要体现在系统、信号、屏幕、硬件、电池、游戏和内存。

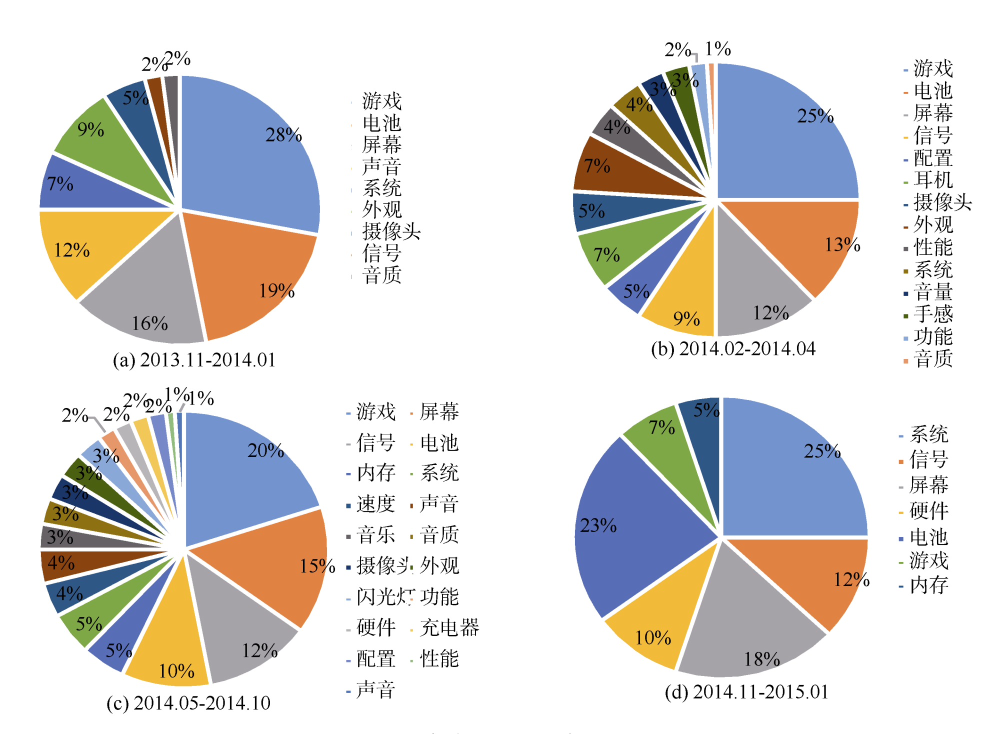


图 6　情感波动因素分析图

从图 5 中可以看出，在 2014 年 2 月–2014 年 4 月这段时间内消费者的情感波动相对平稳，但从 2014 年 5 月–2014 年 10 月这段时间内消费者的负面情感比例不断上升。通过对比 2014 年 2 月–2014 年 4 月和 2014 年 5 月–2014 年 10 月这两个时间段内的饼图可知，消费者在这段时间内对产品特征抱怨的数量在不断增加，从 2014 年 2 月–2014 年 4 月期间的 14 个特征到 2014 年 5 月–2014 年 10 月期间的 19 个特征，增加的特征即是消费者情感波动产生的主要因素。作为生产厂商在发现消费者负面情感比例有不断上升的趋势时，就应该作出相应的销售策略和售后服务。

在不同的时间段内消费者所关注的产品的特征是不同的，但总体可以看出在这 4 个时间段内顾客所关注的产品特征主要集中在游戏、屏幕、信号、电池、系统这几个方面，这几个方面在 4 个时间段内所占的比例一直相对较高。对于手机厂商来说在对下一代产品进行设计时可以根据消费者对这些产品特征的关注情况进行针对性的改进。

为进一步分析消费者对情感波动因素不同维度的评价，通过构建共词矩阵，运用社会网络分析工具 Ucinet 对共现矩阵进行可视化分析，由于网络图过大，因此只展示其中一部分，如图 7 所示：

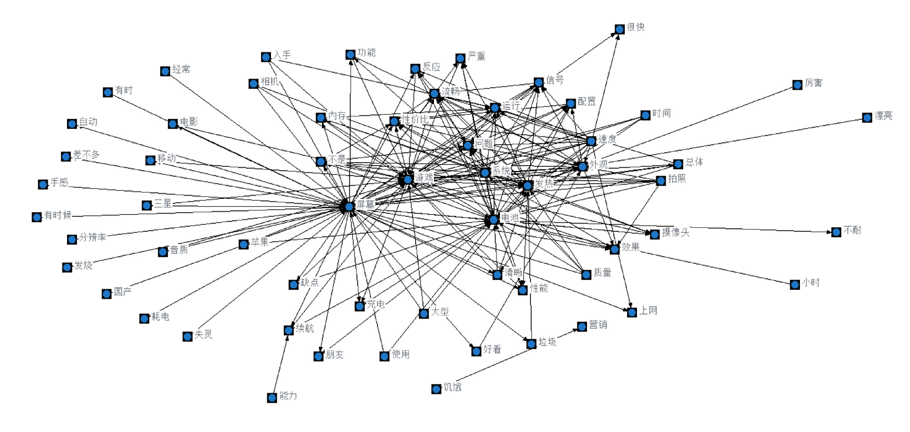


图 7　共词矩阵可视化分析

通过图 7 可以看出屏幕这个产品特征词语和失灵、发烧、手感等一些词语相关联，还有游戏和反应、流畅等一些词语相关联，通过网络图还可以看出和其他一些产品特征词相关联的词语。通过共词矩阵的可视化分析，生产厂家很容易看出消费者对产品特征哪些维度的评价，进一步为生产厂家做出决策提供参考。

5 结　语

本文基于产品评论信息构建消费者情感波动分析模型，将产品评论挖掘技术和情感分析技术应用于顾客对产品的情感波动分析中，并以某手机为例，进行实证研究。研究结果表明，本文构建的模型能够有效地分析消费者对产品的情感波动情况以及产生情感波动的主要原因，相比其他学者所关注的维度，该模型更加宽泛、准确地找出消费者的主要关注点，同时该研究结果与人们购买手机的观念具有一致性，对企业做出相应的决策具有重要的现实意义。今后研究的重点主要是更精确地提取出产品的特征词，在计算情感词的极性值算法上进一步改进，提高计算情感词极性值的准确度，同时考虑其他维度信息对消费者情感波动产生的影响。

参考文献：

[1] Xu K Q, Liao S S Y, Li J X, et al. Mining Comparative Opinions from Customer Reviews for Competitive Intelligence [J]. Decision Support Systems, 2011, 50(4): 743-754.

[2] Rodrigues R G, Dores R M D, Camilo-Junior C G, et al. SentiHealth-Cancer: A Sentiment Anaiysis Tool to Help Detecting Mood of Patients in Online Social Networks [J]. International Journal of Medical Informatics, 2016, 85 (1): 80-95.

[3] Salehan M, Kim D J. Predicting the Performance of Online Consumer Reviews: A Sentiment Mining Approach to Big Data Analytic [J]. Decision Support Systems, 2016, 81: 30-40.

[4] Deng X, Chen R. Sentiment Analysis Based Online Restaurants Fake Reviews Hype Detection [C]. In: Proceedings of APWeb 2014: Web Technologies and Applications. Springer International Publishing, 2014: 1-10.

[5] Wallace B C, Paul M J, Sarkar U, et al. A Large-scale Quantitative Analysis of Latent Factors and Sentiment in Online Doctor Reviews[J].Journal of the American Medical Informatics Association, 2014, 21(6): 1098-1103.

[6] 翟东升，徐颖，黄鲁成，等. 基于产品评论挖掘的竞争产品优势分析[J]. 情报杂志，2013, 32(2): 45-51. (Zhai Donsheng, Xu Ying, Huang Lucheng, et al. The Advantage Analysis of Competitive Product Based on Product Reviews Mining [J]. Journal of Intelligence, 2013, 32(2): 45-51.)

[7] 施国良，石桥峰. 基于文本挖掘的不同购物网站商品评论一致性研究[J]. 现代图书情报技术，2011(12): 64-68. (Shi Guoliang, Shi Qiaofeng. Text Mining-based Consistency of Product Reviews in Different Shopping Websites [J]. New Technology of Library and Information Service, 2011(12): 64-68.)

[8] 王伟，王洪伟. 特征观点对购买意愿的影响：在线评论的情感分析方法[J]. 系统工程理论与实践，2016, 36 (1): 63-76. (Wang Wei, Wang Hongwei. The Influence of Aspect-based Opinions on User’s Purchase Intention Using Sentiment Analysis of Online Reviews [J]. Systems Engineering-Theory & Practice, 2016, 36 (1): 63-76.)

[9] 吴丽华，冯建平，曹均阔. 中文网络评论的 IT 产品特征挖掘及情感倾向性分析[J]. 计算机与数字工程，2012, 40(11): 52-55. (Wu Lihua, Feng Jianping, Cao Junkuo. Mining Information Technology Product Features and Sentiment Orientation from Chinese Customer Reviews [J]. Computer & Digital Engineering, 2012, 40 (11): 52-55.)

[10] 田雪筠.面向用户评论的企业竞争情报挖掘研究[J]. 图书馆学研究，2014(19): 71-78. (Tian Xueyun. Research on Enterprise Competitive Intelligence Mining Based on User Comments [J]. Library Science Research, 2014(19): 71-78.)

[11] 吴丽云，陈芳英. 基于网络评论内容分析的餐饮消费者行为研究[J]. 人文地理，2015, 30(5): 147-152. (Wu Liyun, Chen Fangying. Research on Catering Consumer Behavior Based on the Content Analysis of Network Evaluation [J]. Human Geography, 2015, 30 (5): 147-152.)

[12] 黄卫东，陈凌云，吴美蓉.网络舆情话题情感演化研究[J]. 情报杂志，2014, 33(1): 102-106. (Huang Weidong, Chen Lingyun, Wu Meirong. Research on the Sentiment Evolution of Online Public Opinion Topic [J]. Journal of Intelligence, 2014, 33 (1): 102-106.)

[13] 倪瑜泽，彭蓉，孙栋，等. 基于用户评论的潜在演化需求发现方法[J]. 武汉大学学报：理工版，2015, 61 (4): 347-355. (Ni Yuze, Peng Rong, Song Dong, et al. Potential Evolution Requirements Detect Based on User Comments [J]. Journal of Wuhan University: Natural Science Edtion, 2015, 61(4): 347-355.)

[14] 何天翔，张晖，李波，等. 一种基于情感分析的网络舆情演化分析方法[J]. 软件导刊，2015, 14 (5): 130-134. (He Tianxiang, Zhang Hui, Li Bo, el at. An Analysis Method of Network Public Opinion Evolution Based on Emotion

Analysis[J]. Software Guide, 2015, 14 (5): 130-134.)

[15] 李超雄，黄发良，温肖谦，等. 基于动态主题情感混合模型的微博主题情感演化分析方法[J].计算机应用，2015, 35(10): 2905-2910. (Li Chaoxiong, Huang Faliang, Wen Xiaoqian, et al. Evolution Analysis Method of Microblog Topic-sentiment Based on Dynamic Topic Sentiment Combining Model [J]. Journal of Computer Applications, 2015, 35(10): 2905-2910.)

[16] R 语言实现中文分词[EB/OL]. [2016-06-20]. http://blog.sina.com.cn/s/blog_631d4d990102vcjj.html. (R Language to Achieve the Chinese Word Segmentation [EB/OL]. [2016-06-20]. http://blog.sina.com.cn/s/blog_631d4d990102vcjj.html.)

[17] 梁喜涛，顾磊. 中文分词与词性标注研究[J]. 计算机技术与发展，2015, 25(2): 175-180. (Liang Xitao, Gu Lei. Study on Word Segmentation and Part-of-Speech Tagging [J]. Computer Technology and Development, 2015, 25(2): 175-180.)

[18] 情感分析用词语集(beta 版)[EB/OL]. [2016-06-22]. http://www.keenage.com/html/c_bulletin_2007.htm. (Analysis of Emotional Words Set (beta) [EB/OL]. [2016-06-22]. http://www.keenage.com/html/c_bulletin_2007.htm.)

[19] Turney P D, Littman M L. Measuring Praise and Criticism: Inference of Semantic Orientation from Association [J]. ACM Transactions on Information System, 2003, 21(4): 315-346.

[20] 王晓耘，史玲玲. 基于网络评论情感量化的商品综合评分模型[J]. 杭州电子科技大学学报：社会科学版，2016, 12(3): 8-15. (Wang Xiaoyun, Shi Lingling. A Comprehensive Scoring Model of Products Based on Emotion Quantification of Web Reviews [J]. Journal of Hangzhou Dianzi University: Social Science Edition, 2016, 12(3): 8-15.)

[21] 魏慧玲. 文本情感分析在产品评论中的应用研究[D]. 北京：北京交通大学，2014. (Wei Huiling. Application Research of Text Sentiment Analysis in Product Reviews [D]. Beijing: Beijing Jiaotong University, 2014.)

[22] 单晓红，杨柳. 网络产品评论挖掘研究[J]. 计算机系统应用，2014, 23(2): 1-6. (Shan Xiaohong, Yang Liu. Research on Online Product Review Mining [J]. Computer Systems & Applicantions, 2014, 23(2): 1-6.)

作者贡献声明:

林园园：提出研究思路和实验方案；
张凡，李长江：进行实验；
林园园：采集、清洗和分析数据；
林园园：论文起草；
战洪飞，余军合：论文修改和最终版本修订。

利益冲突声明:

所有作者声明不存在利益冲突关系。

支撑数据:

支撑数据由作者自存储，E-mail: 1051881967@qq.com。

[1] 林园园，战洪飞，余军合，李长江，张凡. comment.txt. 手机产品评论数据.

收稿日期: 2016-06-29
收修改稿日期: 2016-08-04

Using Product Reviews to Analyze Sentiment Fluctuation of Consumer

Lin Yuanyuan　Zhan Hongfei　Yu Junhe　Li Changjiang　Zhang Fan
(The Faculty of Mechanical Engineering and Mechanics, Ningbo University, Ningbo 315211, China)

Abstract: [**Objective**] This paper establishes a model to analyze the sentiment fluctuation of consumers with online product reviews. [**Methods**] We constructed the model with product review mining and sentiment analysis techniques. And also examined the influence of conjunctions to sentence sentimental tendentiousness and then calculated their weights. [**Results**] The proposed model effectively analysed online reviews of one mobile phone posted on Jingdong and Zhongguancun Online from November 2013 to January 2015. [**Limitations**] Only included the total number and frequency of product feature keywords from reviews posted in neighboring time slots. [**Conclusions**] The proposed model could effectively analyze the developing trends and reasons of consumer sentiment fluctuation over a period of time, which provides valuable information to enterprise decision making.

Keywords: Data analysis　Text mining　Emotional fluctuation analysis model